

**Economics and Organization of Bibliographic Data—the A&I Perspective**  
**Linda Beebe, American Psychological Association**

Good morning. I'm pleased to be with you today as you consider the important question of how we can manage and provide access to the explosion of knowledge in today's world and the related question of how we can best use technology to accomplish the tasks. From your work to date, I see that the Working Group is also looking at ways we can integrate all the multiple standards and protocols that exist today. In my mind, that is a crucial component of the task.

As a non-cataloger, I lack the specialized knowledge that many of you share. However, I bring the perspective of the abstracting and indexing world—and our entire purpose for being is to aid discovery and finding of the scholarly, scientific, and technical literature. A&Is are very good at critical metadata, and most of us are very good at controlled vocabulary, another essential ingredient in managing and providing access. Over the past several years, we have evolved into linking tools, so that users can get directly from an A&I service to the full text. Researchers can use the breadth and precision of the bibliographic database to search the full range of information on a subject area, focus on what they really need to answer their questions, then link to the full text.

That is, they can if we work in collaborative ways to integrate our systems. There is so much inter-related activity in primary publication, bibliographic publication, library resource management, and search engine technology that no one can afford to function as a silo. I see some evidence that integration is occurring, even though the worlds remain somewhat separate. Two quick examples: First, libraries are working with search engines, most notably Google, to connect to their link resolvers to get the users to content they have purchased access for. Google offers breadth, but not precision or additional limiters so essential to the serious scholar; when the secondary service is added into the mix, there is great enhancement of the results. Second, CrossRef has achieved a fair amount of success with multiple resolution to solve the "appropriate copy" question, which is a key issue for those of us who distribute our content through multiple vendors and for the customers who buy the content. Key elements in the success of CrossRef are broad collaboration, persistent identification with Digital Object Identifiers and standardized metadata. These things enable seamless linking across disparate services. So, to me, an essential task is integration—and consequently dismantling silos.

As you know, the explosion in knowledge is now outstripping the ability of even the most skilled researcher to keep up, even though the sciences become ever more specialized. Yet, researchers must sift through the vast array of choices to discover the right information. I understand that one of the questions the Group is wrestling with around economics is "Is some access better than none?" In some cases, that may be true. But when researchers are dealing with complex questions that affect human lives, a little knowledge can truly be dangerous. Missing an important finding could result in a critical error that could cause physical harm to people or cost a researcher wasted time following a path that has already proved to be fruitless. I believe that descriptive practices must support enabling the researcher to search across the vast body of literature and find what they need—seamlessly and quickly.

There is no question that we need standards or we will have chaos. The question is how many standards, can they be simplified in this complex world, and can they, too, be integrated? In the publishing world, we have learned that our schemas need not be as complicated as we first thought—when we began tagging, we tagged everything because we could and thought we needed

July 9, 2007

to. That led to constant change and little potential for integration—the DTD as house style did not serve anyone well. We are now refining schemas. Perhaps simplification in other arenas might help us solve our access problems across the literature as well. It seems to me that we will achieve good interoperability as discussed in the May meeting of the Working Group only by examining the tasks and meeting points of linked systems—and we know from re-engineering projects that this examination often leads to simplification in workflow and other processes as well.

At the same time, the growing variety of multimedia content is creating new issues in our management of access to content. That complexity exists for both the producer and the provider of access.

In various discussions of this subject, people have noted the librarians' need to manage the resources and also to bring them to the attention of the end users in the most effective way; further, the existing tools are not always optimal for both purposes. We publishers—primary and secondary—face a similar dilemma. We have two audiences: the librarian who is our customer and the end user. The librarian has legitimate needs for sophisticated bells and whistles; the user wants the impact of those bells and whistles, but they want them invisible. An example would be use of the controlled vocabulary. The librarian views a good thesaurus, with its hierarchy, scope notes, history notes, and so on, as a key component of good metadata. The user simply wants synonym searching, even if they don't know that's what they are looking for. They want to key in one term and get all the "more like this." So, producers must answer both needs.

Another economic challenge to both librarians and publishers is educating the user about what they can do. We've had to discard the notion that it takes very specialized knowledge to find information. Eliminating complex search protocols and making it easier to find what you are looking for is *not* dumbing down the literature—and all of us need to keep making it easier and easier. Even with the most easy-to-use systems, however, we will have to get the user's attention. In talking with graduate students, we've been startled by two competing facts—they have enormous time pressures; at the same time, they often work in the least efficient ways. Although they could link to a full-text article instantly, they copy and paste information into a new search. Shown that they can get articles in a specific methodology in an instant without wading through irrelevant content *if* they use the right field, they are ecstatic, but they would not look for that field on their own. We must make the learning curve as gently sloping as possible—but we also need to factor in some education. This is another area in which publishers and librarians can work together.

Overall, it seems to me that the crucial tasks—integration and the elimination of silos, finding ways to search across all information and focus on the most important, creating interoperability, serving the needs of librarians for managing the information and users for finding it, and reducing the costs—can be accomplished only by the key parties in this endeavor working together. The evolution of electronic publishing, while it has introduced more complexity, has made it possible for publishers to assist librarians in managing content in ways that were not previously possible. There is no question that we can use the Internet more effectively than we do today, but we cannot leave the "how" of that to any one of the several interested parties.

Thank you for allowing me to represent the abstracting and indexing industry in your discussions today. The board of NFAIS is very interested in your deliberations, and I'll be certain to convey what I have learned today.

July 9, 2007